Measuring the Power of Learning.®

Research Report

ETS RR–16-33

An Evaluation of the Usefulness of Prosodic and Lexical Cues for Understanding Synthesized Speech of Mathematics

Lois Frankel

Beth Brownstein

October 2016

ETS Research Report Series

Since its 1947 founding, ETS has conducted and disseminated scientific research to support its products and services, and to advance the measurement and education fields. In keeping with these goals, ETS is committed to making its research freely available to the professional community and to the general public. Published accounts of ETS research, including papers in the ETS Research Report series, undergo a formal peer-review process by ETS staff to ensure that they meet established scientific and professional standards. All such ETS-conducted peer reviews are in addition to any reviews that outside organizations may provide as part of their own publication processes. Peer review notwithstanding, the positions expressed in the ETS Research Report series and other published accounts of ETS research are those of the authors and not necessarily those of the Officers and Trustees of Educational Testing Service.

The Daniel Eignor Editorship is named in honor of Dr. Daniel R. Eignor, who from 2001 until 2011 served the Research and Development division as Editor for the ETS Research Report series. The Eignor Editorship has been created to recognize the pivotal leadership role that Dr. Eignor played in the research publication process at ETS.

ETS Research Report Series ISSN 2330-8516

RESEARCH REPORT

An Evaluation of the Usefulness of Prosodic and Lexical Cues for Understanding Synthesized Speech of Mathematics

Lois Frankel & Beth Brownstein

Educational Testing Service, Princeton, NJ

The work described in this report is the second phase of a project to provide easy-to-use tools for authoring and rendering secondary-school algebra-level math expressions in synthesized speech that is useful for students with blindness or low vision. This report describes the development and results of the second feedback study performed for our project, Expanding Audio Access to Mathematics Expressions by Students With Visual Impairments via MathML. That study focused on the use of certain prosodic and lexical elements in the ClearSpeak speech style and served as a basis for further refinements in that style's definition and implementation in the MathPlayer software. The primary parameters evaluated are students' success in drawing conclusions about the content and structure of certain math expressions and their perceptions regarding the helpfulness of the pace and wording of different text-to-speech renditions of the same or similar mathematical expressions. Please see Appendix A for information on obtaining a version of this report that is fully accessible using the tools described.

Keywords math; accessibility; blindness; visual impairment; text-to-speech; MathML; prosody; algebra; STEM; ClearSpeak; assistive technology; screen reader

doi:10.1002/ets2.12119

Students with visual impairments (SVIs) are known to have a large gap in math achievement compared to students without disabilities (Blackorby, Chorost, Garza, & Guzman, 2003, Chapter 4). To narrow this achievement gap, it is necessary to address the access gap: Although the *Nemeth Braille Code for Mathematics and Science Notation* (Nemeth, 1972) provides a standard for braille mathematics materials, many SVIs are not proficient in Nemeth Code, and even for those who are proficient, braille materials are often not available when needed. Once technological barriers have been overcome, audio is an alternative that can be used either in addition to braille or when braille materials are not available. One advantage to audio is that it can be made available quickly, particularly for materials that are already available electronically.

Our project, one portion of which is described here, attempts to overcome the technological barriers to meaningful audio access to mathematics.

This report describes the purpose, methodology, and results of the second feedback study performed for our project, Expanding Audio Access to Mathematics Expressions by Students With Visual Impairments via MathML. The project was funded by a U.S. Department of Education, Institute of Education Sciences Special Education Development Grant (R324A110355), which supported the iterative development of the ClearSpeak speech style, authoring tools, interactive navigation, and integration with Microsoft Word. Four feedback studies—one focusing on speech styles, one focusing on the use of certain prosodic and lexical elements in the speech style (described in this report), one on the interactive navigation capability, and one on authoring by teachers and service providers—guided development, culminating in a final pilot in Spring 2015.

Background

For text-only materials, it is relatively easy to prepare accessible electronic materials that SVIs can use with screen readers to achieve audio access. Math expressions present serious challenges to audio presentation. Some expressions are typically presented in print in two dimensions (e.g., fractions, superscripts); some use symbols that are not used in plain text and/or whose meaning in math expressions differs from their meaning in text; some have complex structures that can be difficult to keep track of (e.g., nested parentheses, square roots of complicated expressions, complicated exponents, fractions within

Corresponding author: L. Frankel, E-mail: lfrankel@ets.org

fractions). Until our project and others with similar goals[1] leveraged the MathML markup language, these challenges had prevented screen readers from providing meaningful access to mathematical expressions. These issues are discussed in more detail in our previous report (Frankel, Brownstein, Soiffer, & Hansen, 2016).

Our project focuses on improvements to text-to-speech (TTS) renditions of math expressions. Audio access to mathematics can, in principle, make use of nontext sounds as well; however, nontext sounds cannot be integrated with currently available screen reading software, whereas TTS renditions can be. The improvements we have made to TTS come under two general headings: (a) the development of classroom-like speech (the ClearSpeak style), which includes two essential features: speech for the math that is correct, familiar, and understandable, and prosody (e.g., pausing) that is similar to that found in human speech in the classroom; and (b) interactive navigation. We previously reported (Frankel et al., 2016) on the results of our initial tests of the ClearSpeak style for synthetic speech of algebraic expressions. Those initial tests were conducted on the ClearSpeak prototype, with its main focus on correct, familiar, and understandable speech for the math. Built into that prototype was some very basic prosody, in the form of natural pauses separating mathematical elements. In our first report, we described how the first iteration of ClearSpeak was developed and documented the results of the first feedback study. In that study, students listened to parallel but different math expressions (e.g. clones) in our ClearSpeak (classroom-like) style and in the other two speech styles included in the MathPlayer software. We found that students' subjective ratings of the speech (familiarity, ease of understanding, confidence that they had understood the math expression, and preference for one style over others), along with the objective measure of their success in decoding math expressions, were higher for ClearSpeak than for the other two styles. This first study showed us that we were on a reasonable trajectory in our development of ClearSpeak.

The next step was to consider potential improvements and refinements to the ClearSpeak speech rules. Based on research on how people understand spoken math expressions (see, e.g., Gellenbeck & Stefik, 2009) as well as our own experience in scripting math expressions so they can be spoken, we saw a particular need to provide speech that helps the listener discern an expression's structure (including boundaries for fractions, exponents, roots, parentheses, etc.) and any included (nested) substructures while minimizing memory load to the extent possible. While lexical cues ("end root," "close paren," "end fraction") are least susceptible to ambiguity, they add to a listener's memory load and may thus contribute to the very confusion their inclusion attempts to alleviate. Accordingly, we were interested in determining whether certain prosodic cues could take the place of at least some lexical cues. To guide our decisions in that regard, our second study, conducted in late Spring 2013, investigated the usefulness (as measured by successful recognition) and perceived helpfulness of various prosodic and lexical cues for improving recognition of boundaries of such structures as fractions and square roots and for clarifying information provided by parenthetical groupings, including nested parentheses. We focused our investigation in the second study on those structures because of their importance in the secondary-school algebra curriculum and because they exemplify common accessibility barriers found in a range of mathematical structures. This report describes that second study and its results.

Prosody

Prosody in synthetic speech can include any or all of the following:

- Adjustments to speech rate
- Adjustments to pitch
- Adjustments to volume
- Pauses of varying lengths

Prosody does not include the use of any nonspeech sounds, such as *earcons* (Stevens, Edwards, & Harling, 1997), or stereo or surround-sound (*spatializing*) to alter the perceived location of a sound. An additional type of adjustment that might be made, and which may or may not be construed as prosodic, is specifying different TTS voices to be used for different portions of expressions or nesting levels or to convey other types of structural information.

The use of prosody for improving TTS renditions of mathematics expressions was pioneered in the 1990s by T. V. Raman with his ASTER system (Raman, 1994), which used such prosodic cues as lower or higher pitch to indicate subscripts or superscripts, along with various other prosodic cues. With their AudioMath tool (which, like MathPlayer, works with expressions presented in MathML), Ferreira and Freitas (2005) explored the use of short and long pauses to signal hierarchies in expressions (basing pause length and use of rising or falling tones on patterns found

in human speech). They concluded that prosody in math required further study and that navigation functionality is necessary.

Karshmer, Gupta, and Pontelli (2007) described the benefits and liabilities of using lexical indicators (such as “begin square root”/“end square root”), noting that such indicators are highly effective for disambiguation but add burdens to working memory (p. 13). They also described a variety of approaches to prosody and other types of audio cues.

Gellenbeck and Stefik (2009) found that pauses were very useful for disambiguating the algebraic expressions that they tested but cautioned that additional strategies would likely be needed for some expression types that were outside the scope of their study. The participants in their study were college juniors and seniors majoring in computer science. No mention is made of whether any of the participants had visual or any other disabilities, but because the study asked participants to match spoken algebraic expressions to corresponding printed versions, it must be assumed that all participants had sufficient usable vision to perform this task. They also noted that their target audience was college students with learning disabilities. Some participants listened to math spoken with pauses; others listened to the speech without pauses. They were then asked to rate how well they thought a supplied printed expression matched the audio. As expected, Gellenbeck and Stefik (2009) found that adding pauses resulted in significant improvements in their participants’ ability to disambiguate expressions. They did not investigate the use of pauses of different lengths.

Bates and Fitzpatrick (2010) described the advantages and disadvantages of lexical and prosodic cues, spatialization (particularly left-right localization of start- and end-sounds for fractions and other structures), and nonspeech sounds. They proposed a model that combines these cues with *spearcons* (sped-up TTS) with the aim of achieving improved comprehension while minimizing the cognitive “overhead” required to process the various types of auditory information. Similarly to the other researchers mentioned, they noted that processing mathematical information via audio is inherently more cognitively intensive than is doing so via vision, since visual material serves as “external memory” (p. 408) for sighted individuals. They concluded that goals for systems for audio renditions of mathematics should include resolving ambiguities, maximizing cognitive efficiency, and “temporal control over the material in the form of browsing and overview capabilities” (p. 413) — that is, some sort of interactive navigation.

Although each of the strategies mentioned previously has some promise for improving audio accessibility for math, none of them is readily integrated with screen readers or other commonly used assistive technology; rather, they are implemented as stand-alone systems. Our project is the first to enable the integration of prosodic cues into screen reader output from documents in Microsoft Word or browsers, as opposed to premade audio files or self-contained software environments.

Method

After identifying prosodic elements that are typically used in classroom speech for disambiguating boundaries and determining which of those were most universally and consistently supported by the synthetic speech engines available, we reduced our prosodic tool kit to pauses and rate adjustments. The prosodic elements not used, such as changes in pitch or volume, though technically supported by ClearSpeak, were either not supported or supported inconsistently by many of the speech engines or voices in current use, and so were not included in the ClearSpeak rules. Nonspeech sounds and spatializing were also not supported in our software environment and so could not be explored.

We devised the second feedback study to focus on three specific research questions revolving around certain types of prosodic and lexical cues that could be supported reliably in our environment. The research questions focus on structures that (a) are important in the early algebra curriculum and (b) exemplify the types of boundary-identification issues and related accessibility barriers typically found in mathematical structures. It was necessary to restrict the focus to a few types of structures in order to keep the amount of time required of the students to a manageable level. The three research questions follow.

1. For differentiating boundaries of square roots, how do two prosodic methods (extended pauses at the end of the square root and increased speech rate for the portion of the expression under the square root) and one lexical (“end root”) method compare with regard to students’ success in identifying the expressions and students’ indications of favorability toward each method?

2. For clarifying nesting levels of multiple sets of parentheses, how do the prosodic cues of uniform or graduated-length pauses between different nesting levels, with or without lexical cues indicating such levels, compare with regard to students' indications of favorability toward each method?
3. For expressions involving multiplication of parenthesized expressions, how does speaking the implied "times" and speaking (or not speaking) the parentheses affect students' success at identifying the expressions and their indications of favorability toward the speech?

In the section "Results by Research Question," we discuss each research question in detail along with the items created to address the question and the results of our inquiry.

Participants

Participants were 22 students with visual disabilities (blindness or low vision). Students were aged 14–19 and were in grades 8–12. Ten students were blind, and 12 had low vision. Seven were enrolled in inclusive settings in Kentucky; the remainder were enrolled in schools for the blind in Texas and Washington states. Each participating student was taking or had completed Algebra 1, was fluent in English, and did not have a significant cognitive disability.

Sampling Procedures

Participants were recruited through two participating schools for the blind and through a consultant who recruited students being educated in inclusive settings. Students were given $25 in gift cards for completing the study. The internal review board approval and signed informed consent forms were obtained prior to data collection.

Research Design

Instruments

The study used three instruments: a student background questionnaire, a student background questionnaire for teachers, and a math instrument with feedback questions.

The student background questionnaire asked students to self-report their math- and vision-related background and their history with using various forms of assistive technology for math.

The student background questionnaire for teachers was parallel to the questionnaire for students and asked teachers the same questions (on behalf of their students) that the students had answered for themselves. Central topics of the two questionnaires included the following:

- Student's current best mode for accessing math (print, braille, read aloud, various types of assistive technology)
- Perceived usefulness for the student of the various modes for accessing math (print, braille, read aloud, various types of assistive technology)
- Student's proficiency with fractions, square roots, and parentheses (grouping)
- The degree to which the student uses visual versus nonvisual methods of accessing math

The math instrument with feedback questions was administered to students. They completed a practice version of the instrument, which included introductory information about the software and procedure, to provide them with some familiarity with the software. Then they completed one three-section form in which they used the Window-Eyes screen reader to read text and math, answer questions about the math, and answer feedback questions about the way the math was spoken.

Study Manipulation

Students were randomly selected to receive one of three versions of the math instrument. The versions differed only in the order in which the various treatments of the math expressions in each section were presented. The reorderings were made at random, and were intended to compensate for any order-effects that might arise, such as a preference for the speech treatment heard first or last, or any learning effects from one hearing of a similar math expression to the next.

Delivery Method

The textual portions of the practice and math instruments were provided in a large font (24-point Verdana). As we explained to the participants in introductory material, we left the math expressions in the smaller font (12-point Times Roman) to encourage students to focus on how the math was spoken, not on how it looked. We further reassured the students that this differential formatting was for research purposes only and that when these speaking tools were finished, the math could be formatted at any size that is useful and would usually display at the same font size as the surrounding text. The practice and math instruments were administered on computers on which the study software (Microsoft Word, MathType, MathPlayer, and a version of Window-Eyes with MathPlayer support) had been installed.

Session Procedure

1. Students completed the background questionnaires with the help of the study administrator, who also collected parallel information from the students' math and VI teachers.
2. Administrators provided the students with the practice version of the instrument: a shortened version of the math instrument, using easier math questions. Its purpose was to give students familiarity with the structure of the study instrument and with using Window-Eyes to listen to math from inside Microsoft Word.
3. Administrators provided the students with the correct instrument, based on the random order assignment. The beginning of the instrument explained the purpose of the study and described the procedure to the students.
4. Administrators guided the students through the three sections of the instrument, including the math questions, the feedback questions, the end-of-section questions, and the end-of-study questions. Students proceeded through the instrument at their own pace and were allowed to listen to the math expressions as many or few times as they wished. They were allowed to take breaks as needed and to use any equipment they normally used when studying math, such as braille or paper note-taking equipment, screen enlargement software, and calculators. As needed, administrators scribed responses for those students who did not enter their own responses into the Word document.

Qualifications of Study Administrators

Study administrators were either project consultants who worked at one of the cooperating schools or cooperating school personnel. For students who took notes in braille, the study administrator transcribed those notes into print so that we could consider them when we analyzed the data.

Data Analysis

All responses were entered into a Microsoft Access database. Where appropriate, scores for responses were calculated. Responses to math questions that had a single correct answer (Section 1) were scored *correct* (1) or *incorrect* (0). Responses to math questions where students were to select "all that apply" (Section 3) were scored from 0 to 1, depending on what fraction of the answer choices were correctly selected or correctly not selected. Section 2 did not have scorable math questions. Feedback questions on how certain students were of their answers and how helpful or unhelpful the speech was, were scored from 0 (*just guessed answer/speech was not at all helpful*) to 3 (*very sure of answer/speech was very helpful*). The database and explanations of the scoring and the comparisons to be made in the analysis were shared with a data analyst, who conducted independent samples and chi-square tests. Details are described by research question.

Development of the Items

With consultation from two of our expert advisors—Susan Osterhaus and Maylene Bird (both of whom teach math to students who are blind or visually impaired)—we crafted items to test different prosodic and lexical cues to address each research question. The development involved careful consideration of and experimentation with the details of each treatment to be tested (e.g., how long the pauses should be, by how much the speech rate should

vary) and on the types and complexity of the math expressions to be used. In the resulting instrument, each of the three sections addressed a different one of the three research questions. The next section describes each section of the instrument in the context of the research question it was developed to answer and details the results that were obtained.

Results by Research Question

The three research questions are listed in the "Method" section. This section describes the quantitative and qualitative results for each question. One section of the math instrument was devoted to each question.

Research Question 1 (Boundaries)

Consider the expressions $1 + \sqrt{x+y} - 7$ and $1 + \sqrt{x+y-7}$. When one listens to the expression spoken aloud as "one plus the square root of x plus y minus seven" without any prosodic or additional lexical cues, the quantity under the square root could appear to be x, $x+y$, or $x+y-7$. However, saying "end root" to disambiguate the expression adds to the memory load. Similar issues arise with regard to the boundaries of fractions, exponents, and other types of expressions. As the guiding principle of ClearSpeak is to emulate, to the extent possible, classroom speech, we hypothesized that because varied pause length and changes to speech speed are often used in human speech to help indicate boundary locations, inserting a long pause at the end of the square root's scope or slightly speeding up the speech within the scope of the square root might be acceptable alternatives to saying "end root." Accordingly, we applied three treatments (extended pause, speedup, and "end root" language) to each of the two expressions, presented them to students in one of three randomized orders, and asked students, after listening to each of the resulting six cases, to answer the following math question:

> The expression under the square root is:
>
> a. x
> b. x plus y
> c. x plus y minus 7
> d. I can't tell

Then we followed up with feedback questions about their level of confidence in their choice and an open-ended free-response question about how the speech helped or hindered their ability to answer the math parsing question:

> How sure are you of your choice:
>
> a. Very sure
> b. Somewhat sure
> c. Somewhat unsure
> d. Just guessed (use this answer if your answer to the math question was that you couldn't tell)
>
> How helpful for telling what was under the square root were the wording and the way it was paced?
>
> a. Very helpful
> b. Somewhat helpful
> c. Not very helpful
> d. Not at all helpful
>
> What about the way the math statement was worded and paced helped you tell what was under the square root or made it difficult for you to tell what was under the square root?

Table 1 Question 1 Treatments and Results ($n = 22$)

Treatment	Expression	Percent of students correct	Number of students selecting each response: Can't tell	$x+y-7$	$x+y$	x	Confidence (Scale: 0–3)	Helpfulness (Scale: 0–3)
Extended pauses[a]	$1+\sqrt{x+y}-7$ (1A)	50	4	5	11 (correct)	2	Ave: 2.32 *SD*: .945	Ave: 1.82 *SD*: .907
	$1+\sqrt{x+y-7}$ (1B)	68	3	15 (correct)	2	2	Ave: 2.45 *SD*: .91	Ave: 2.09 *SD*: .92
Normal pauses + speedup under root[b]	$1+\sqrt{x+y}-7$ (1C)	59	3	6	13 (correct)	0	Ave: 2.14 *SD*: 1.08	Ave: 1.95 *SD*: .95
	$1+\sqrt{x+y-7}$ (1D)	86	0	19 (correct)	3	0	Ave: 2.59 *SD*: .666	Ave: 2.14 *SD*: .834
"End root"[c]	$1+\sqrt{x+y}-7$ (1E)	82	1	3	18 (correct)	0	Ave: 2.68 *SD*: .716	Ave: 2.55 *SD*: .963
	$1+\sqrt{x+y-7}$ (1F)	95	1	21 (correct)	0	0	Ave: 2.82 *SD*: .664	Ave: 2.68 *SD*: .568

[a]The speech for Expression 1A was: "1 [pause 400] plus [pause 100] the square root of [pause 300] x [pause 25] plus [pause 25] y [pause 2750] minus [pause 100] 7." Note particularly the very long (2750 msec) pause at the square root boundary. In Expression 1B, the speech was "1 [pause 400] plus [pause 100] the square root of [pause 300] x [pause 25] plus [pause 25] y [pause 25] minus [pause 25] 7 [pause 2750]." Again, there is a 2750 msec pause at the square root boundary, which coincides with the expression's boundary.

[b]The speech for Expression 1C was: "1 [pause 400] plus [pause 100] the square root of [pause 300] [rate 125] x [pause 25] plus [pause 25] y [endrate] [pause 2750] minus [pause 100] 7." Notice the 125% speech rate inside the square root and the 2750 msec pause at the square root boundary. In Expression 1D, the speech was "1 [pause 400] plus [pause 100] the square root of [pause 300] [rate 125] x [pause 25] plus [pause 25] y [pause 25] minus [pause 25] 7 [endrate] [pause 2750]." Again, there is a 2750 msec pause at the square root boundary, which coincides with the expression's boundary, and the 125% speech rate inside the square root.

[c]For Expressions 1E and 1F, respectively, no adjustments were made to pauses or speech rates, and the speech was set (by the predefined rules and preferences) to say "end root" at the close of the square root.

Question 1 Results

Quantitative Analysis

Table 1 shows how many students selected each answer choice for the math question as well as the mean and standard deviations of their indications of confidence level ("how sure are you of your choice?") and of the helpfulness of the pace and wording of the expression. Students heard the expressions and their treatments in varied orders, as previously mentioned.

Regardless of treatment, more students accurately identified the scope of the radical when the radical ended at the end of the expression (1B, 1D, and 1F) than did so when the expression continued following the close of the radical (1A, 1C, and 1E). Chi-square tests found that these differences were significant for the normal pauses plus speedup treatment (1C vs. 1D, Sig. (2-sided) = .042). They were not significant for the end root treatment (1E vs. 1F, Sig. (2-sided) = .154) or for the extended pauses treatment (1A vs. 1B, Sig. (2-sided) = .220).

Comparing treatments for the two expressions taken together, students most accurately identified the expression under the radical for the "end root" treatment (1E and 1F), followed by the normal pauses and speedup treatment (1C and 1D), and had least success with the extended pauses treatment (1A and 1B). The chi-square test showed that the difference in successful identification of the expression between the "end root" treatment and the other two treatments was significant: End root (1E and 1F) versus normal pauses plus speedups (1C and 1D) was preferred with significance = .017. End root (1E and 1F) versus extended pauses (1A and 1B) was preferred with significance = .005. In sum, neither prosodic treatment succeeded in compensating for the absence of explicit lexical end-markers.

When the questions about students' confidence in their responses were analyzed with a chi-square test, the differences were not found to be significant. Differences in students' responses to the questions about how helpful they found the speech were, based on a chi-square test, significant for the end root treatment (1E and 1F) versus the extended pauses

treatment (1A and 1B) with significance = .007 and the end root treatment versus the normal pauses plus speed-ups treatment (1C and 1D) with significance of .010. In each case, students indicated that they found the end root treatment the most helpful.

Qualitative Findings

Students' responses to the open-ended questions were helpful in comparing our expectations for a given treatment with the students' perceptions. Student responses are quoted nearly verbatim. Words in brackets indicate corrections of typos or misspellings; our own comments are in some cases also inserted in brackets.

Long Pauses to Indicate End of the Square Root's Scope (Expressions 1A and 1B)

A fair bit of polarization of sentiment was present regarding the use of the extended pauses. Our intent was that putting a long pause after the end of the expression under the radical would help a student realize that the next portion of the expression was outside the radical. Some students understood that; others did not. For the expression with only "$x + y$" under the radical (Expression 1A in the instrument), of the students who correctly identified the scope of the radical, two found the wording and pace very helpful, most found it somewhat helpful, and a few found it not very helpful.

Some who found the long pause helpful commented:

- The pause helped me identify when the root was over and what was after it.
- It was helpful to have the really long pause.

Others who, although correctly identifying the expression, did not appreciate the pause, said:

- Another long pause not helpful.
- It was helpful that it read slower and it made it more difficult that it did not say end of square root. It just had a long pause.
- The pauses make it difficult to understand what is happening.

Others had mixed feelings about the pause:

- Didn't know if the pause meant the root ended. But if I knew that the pause meant that the root ended that would be helpful. It wasn't explained what the pause meant before I heard it.
- The square root symbol was spoken well, but it was somewhat difficult to tell where it ended, I assume the pause indicated the end.

The students who indicated they couldn't tell what was under the square root or answered incorrectly tended to express more negative views on the long pauses, and some complained about the lack of end language.

- Well it was difficult to tell what was under the square root because there was an open square root indicator but the voice didn't read the terminate root indicator, therefore, I couldn't tell when the square root ended.
- It didn't say end root but had multiple hesitations. It was very vague whether or not there were other additional symbols under the square root.
- The long pause is not helpful. It wastes time and I have to keep the numbers in my head.

Some, however, still found the pauses helpful, not because of their intended significance, but because they slowed down the speech:

- It was a bit slower and more clear.

One, who thought only "x" was under the square root, misinterpreted the pause:

- I got that it paused the x to show that nothing else would be in the square root.

One apparently didn't notice the pause after "$x + y$" and thought there should have been one:

- I felt it was a bit difficult because there was no pause somewhere in the expression. I mean, after the square root it just speaks on and on. I think if there was a pause after that square root, or maybe a pause before, it would help with organizing the expression.

And one found the treatments difficult to distinguish, or at least did not find any preferable to any others:

- They seem all the same to me. I can tell that they are reading it differently, but they are still very similar for my understanding.

For the expression with "$x + y - 7$" under the square root (Expression 1B in the instrument), most students who correctly identified what was under the square root said they were very or somewhat sure of their answer and found the wording and pacing of the speech very or somewhat helpful. One student found it not very helpful. Because the long pause at the end of the square root was, for this expression, also at the end of the entire expression, fewer students noticed the role of the pause in indicating the end of the root. One who did notice that difference had listened to Expression 1A (with only "$x + y$" under the square root) first, noted the pause at the end of the square root in that case, noticed the absence of a pause in Expression 1B, and inferred that the root ended at the end of the expression: "Because there was no pause I figured that all the expressions were under the square root." That was the student who made the comment quoted earlier regarding Expression 1A, "Didn't know if the pause meant the root ended. But if I knew that the pause meant that the root ended that would be helpful. It wasn't explained what the pause meant before I heard it."

One student (who correctly identified what was under the square root but found the speech not very helpful thought there was no pause and would have wanted one:

- I feel like it just needs a pause, as I said in my previous comment. If there was a pause I think it would help listeners understand what is being read. Just speaking a expression through, may not stick in some heads, and it really didn't help me.

As with Expression 1A, many students were expecting an "end root" statement:

- It was difficult to tell what was under the square root symbol because it did not say terminate symbol.
- I think that the pace at which the problem was read was very good but they forgot to put the terminate root indicator.
- It said square root at the beginning and had the rest under the root because it did not say end root or pause.

And one thought the pauses made it more difficult to follow the expression:

- To tell what was under the square root, I think that the [pace] was at a good point, but I think minimizing the [space] between the plus, minus, variables and numbers under the square root would make it easier to keep the numbers bundle[d] when one tries to remember the statement in the future.

Standard Pauses With a Speed-Up Under the Square Root (Expressions 1C and 1D)

Expression 1C used this treatment for the expression with only "$x + y$" under the square root. For this expression and treatment, the students who correctly identified the square root's scope were of mixed opinions on the value of the pauses and the speedups, which, as seen in Table 1, appeared to do little to increase accuracy over the version (Expression 1A) with only the extended pause at the end of the square root. Those who noticed pauses and speedups commented:

- The pause and the way it was read helped me identify when the root ended and what was after it.
- It was helpful that it said square root before the expression started and paused after the root so that I could tell exactly what was inside the root.
- I didn't like the pause it made once the square root part ended. I felt that it was kind of confusing. It was hard to understand what was part of the square root and what wasn't. I know what they were going for, but I think it would [be] confusing for a blind person to understand.
- I liked the wording of it, but the fact that it sped up and slowed back down made it hard to understand.
- I got it when they paused after the y so I knew that the square root sign had ended.
- The pause made it confusing whether it was the end of the square root or just a pause
- I thought that the way the voice read the problem was very steady and clear, there was a pause before the 7 so that helped me understand the problem better.

Many, including students who correctly identified the expression and those who were incorrect, were adamant in objecting to the absence of an "end root" statement:

- To be completely honest, I found it annoying, how it didn't give the end root signal.

- Everything shouldn't be separated; I need words that put everything under the square root.

As with the pair using extended pauses (Expressions 1A and 1B), within the pair using a speedup under the square root, students found Expression 1D ($1+\sqrt{x+y-7}$) easier to follow than Expression 1C, since in 1D the expression ended at the same place as the square root. There was also an increase in the percent of students accurately identifying the expression under the square root over the version (Expression 1B) that used extended pauses (86% correct vs. 68%).

As with Expression 1C, most who noticed the speedup when listening to Expression 1D were unsure of how they should interpret it:

- I noticed that it started speaking faster when it got to the y in the equation but I [didn't] know what it [meant.]
- Having the change in speed was a little helpful, but still could not tell when the root was ended or if there was anything after it.
- The way it was paced at the end didn't bother me [much], but I can see a [definite] increase in speed.

Some correctly interpreted the (lack of) pauses inside the square root:

- They didn't pause in between the terms so that lead me to believe that $x + y - 7$ was under the square root.

And again, some students requested explicit end language:

- The wording was okay but it was a little confusing because it did not say whether or not the subtraction was under the root symbol.

Use of End Root Language (Expressions 1E and 1F)

Given that so many students expressed the desire for end language, it is not surprising that the students were generally more successful, expressed higher confidence, and found the speech more helpful with this treatment. However, not every student appreciated the end language. For example, for Expression 1E ($1+\sqrt{x+y}-7$), with less under the square root, one student commented "didn't understand when it said end root" and another (who had worked with Expression 1C, which used the speedup treatment, immediately prior to working with Expression 1E) felt that Expression 1E used "extra speech which threw me off at first."

But most did appreciate the end language, for example:

- When it read out the beginning and ending root, it helped organize the expression in my head. It was helpful to hear it read like that.
- Having the end root was very helpful for telling where the root ended and what was after.
- It was helpful that it said square root of and end of square root. It made it difficult when it sped up when it was saying what was under the square root. [We infer that the reference to the speedup means Expressions 1C and/or 1D, both of which this student worked with, in that order, immediately prior to working with 1E.]
- The difference this time was that the voice actually did read the root terminator at the end of the square root which made it much clearer to know when the square root ended. I liked how on the last number there was a pause before it. I believe that perhaps if there were pauses in between the different numbers in the expression, that would make it even easier to comprehend the problem.

For Expression 1F ($1+\sqrt{x+y-7}$), which used the end root treatment and had the end of the square root coinciding with the end of the expression, almost all students correctly identified the scope of the square root. One answered "I can't tell" and commented that they did not understand the term "end root." That student's current or most recent math class was Algebra 1. The student self-described as "just learning" about square roots; the companion teacher questionnaire describes the student's math proficiency level in fractions, square roots, and parentheses as "developing" (the lowest level available in the survey) in each case.

Two students commented that the "end root" statement was not necessary in this case, one specifying that it was needed only when "there is another part of the equations after the square root."

The remaining students generally commented favorably on the use of the end language, for example:

- The terminate root was added back into this math statement. If the software is going to start an expression with a beginning root, then it needs to end it by saying "end root." That is my preference.

Discussion

End language appears to be more helpful than prosody for identifying boundaries of square roots. Although we did not test other structures for end language, we speculate that end language would be similarly helpful or appreciated in other structures, and so we have ensured that such language remains available as an option (it is currently available as a ClearSpeak preference for fractions, square roots, absolute value, matrices, and vectors). Additionally, for the convenience of content creators, we added a single preference setting that allows authors to set end language as the preference for all applicable structures at once.

We tested the various pause and speed-up conditions, which were based on the pauses and speedups typically used by teachers and other live readers for similar expressions, without informing the students of the nature and purpose of the pauses and speedups. We did so in order to see how students would naturally respond to the pauses in the speech, as opposed to creating expectations about them. The drawback, however, of not providing such information is that some of those students who mentioned noticing pauses and changes to the speech rate thought the prosodic elements might be inadvertent, a computer "stumble," or otherwise not intended to convey meaning—even though teachers and other live readers tend to employ similar pauses or speedups when they speak math in the classroom. Students may not have expected such purposeful behavior from a computerized voice. It is possible that providing some explanation of how pauses and other prosodic elements are used, or even the natural process of getting used to computerized math speech (as students may already be used to computerized speech of nonmath material), might improve students' likelihood of benefitting from enhanced prosody. As can be seen in the "Qualitative Findings" section, some students did catch on to the intention behind the pauses, supporting the possibility that further instruction and familiarization would make the prosody more helpful. Our primary goal at this point in the project was to provide speech that students would find understandable with little or no prior instruction, and so it is not surprising that what students most expected (end language) was more understandable to most (but not all) of them. A further inquiry could be conducted as to whether, if students were given opportunities to expect and become familiar with potentially helpful prosodic elements, such as strategic pauses and speedups, they would have different impressions of the elements' usefulness or different relative success in identifying expressions. The results of such an inquiry could provide suggestions for further adjustments to the synthetic speech rules, recommendations for adjustments to human speech guidelines, and additional insights into cognitive processing.

Research Question 2 (Nested Parentheses)

Section 2 focused on the handling of nested parentheses and how different prosodic or lexical treatments might help listeners keep track of the various open/close parentheses pairs in the expression and their relationship to each other (i.e., how they nest). Consider an expression with multiple sets of nested parentheses, such as

$$2\left(\left(x+1\right)\left(x+3\right)-4\left(\left(x-1\right)\left(x+2\right)-3\right)\right).$$

It can be difficult to identify the open/close parentheses pairs, let alone how the various pairs nest inside each other, regardless of how one is accessing the information (visually, tactilely, or auditorily). Sighted readers can indicate the relationship visually, but even this is complicated, as suggested by Figure 1, which uses horizontal braces to mark the nesting levels.

$$2\overbrace{\left(\overbrace{\overbrace{(x+1)}\overbrace{(x+3)}}-4\overbrace{\left(\overbrace{(x-1)}\overbrace{(x+2)}-3\right)}\right)}$$

Figure 1 A visual representation of tracking nested parentheses.

Table 2 Question 2 Treatments

Expression: $2((x+1)(x+3)-4((x-1)(x+2)-3))$	Pauses	Speak parenthesis nesting level (nth language)
Treatment 2A	Uniform	No
Treatment 2B	Uniform	Yes
Treatment 2C	Graduated	No
Treatment 2D	Graduated	Yes

Human speakers typically use pauses and/or visual gestures (when speaking to a sighted audience) to indicate nesting. Multiple levels of nesting, although used in a variety of mathematical structures, is most common (at least in secondary-school algebra) for parentheses. We considered the possibility that longer pauses spoken before and after an outer, or first-level, parenthesis (e.g., the one following the first instance of the numeral 2 and the one at the very end of the expression) and shorter ones before and after innermost, or deepest, parentheses (e.g., the ones around "x – 1" and "$x+2$") might be helpful in tracking the pairs of nested parentheses. (Ferreira & Freitas, 2005, likewise explored the use of pause-length to indicate expression hierarchies.) We also developed a novel lexical cue that enabled the incorporation of speech for the nesting level of each pair of parentheses in the expression ("open/close second paren," "open/close third paren," also called "nth language"), which we thought might be helpful, and tested a combination of the two strategies (graduated pauses plus speaking the nesting level) to see if two potentially helpful strategies could be even more helpful if combined. In total, then, we tested four treatments for this expression, administered in varied orders. Table 2 shows how the two different pause patterns and two different speech patterns were combined in the four treatments.

In this section, there was no math parsing question. That was because the expression was (by necessity, given that we were testing multiply-nested parentheses) so complex that we thought that if we did ask them to attempt such a task, students would just write down the expression as they heard it and parse it from what they had written down rather than from what they had heard. Instead, we asked the students, after they had heard all four treatments of the expression, to rank them in order of helpfulness (and invited them to listen to the expressions again to refresh their memories). Also, after they heard each treatment, we asked them to indicate how helpful they found it and to answer an open-ended feedback question about it:

Suppose you were asked to solve a problem involving this expression, like "What is the value of the expression when x equals a particular number?" or "Simplify the expression by multiplying out and combining like terms." How helpful would it be to have the expression spoken in this way?

a. Very helpful
b. Somewhat helpful
c. Not very helpful
d. Not at all helpful

What about the way the math statement was worded and paced would be helpful or make it more difficult for you to simplify the expression or determine the value of the expression when x equals a particular number?

Question 2 Results

Quantitative Analysis

Table 3 summarizes the ranking and helpfulness scores for the four treatments. The combination of nth language and uniform pauses (Treatment 2B) tended to be more preferred and to be rated as more helpful.

Independent samples tests showed the following.

Concerning use of "open/close nth paren" versus "open/close paren," no significant differences in preference were found for the use of nth paren versus no nth, for the two pause treatments considered together (comparing Treatments 2B and 2D with Treatments 2A and 2C). However, when the comparison is limited to uniform pauses (Treatments 2A vs. 2B), students found the nth language more helpful (mean 1.68 vs. 1.5) than its absence ($p=.006$).

Table 3 Question 2 Results ($n = 22$)

Treatment	2A (uniform pauses, no nth)	2B (uniform pauses; nth)	2C (graduated pauses, no nth)	2D (graduated pauses, nth)
Assigned rank: 1	7	10	3	3
Assigned rank: 2	2	5	7	8
Assigned rank: 3	5	4	7	6
Assigned rank: 4	8	3	5	5
Total	22	22	22	22
Rank (1 – 4; lower number indicates more preferred)	Ave: 2.64 *SD*: 1.293	Ave: 2 *SD*: 1.113	Ave: 2.64 *SD*: 1.002	Ave: 2.59 *SD*: 1.008
Helpfulness (0 – 3; higher number indicates more helpful.)	Ave: 1.5 *SD*: 0.802	Ave: 1.68 *SD*: 1.211	Ave: 1.32 *SD*: 1.086	Ave: 1.55 *SD*: 1.011

Ranking Differences

Concerning whether pauses should be uniform or graduated, independent of use or absence of nth language (Treatments 2A and 2B vs. Treatments 2C and 2D), some significance was found for a preference of uniform over graduated pauses: The uniform pause cases had a mean rank of 2.318 versus graduated pauses, which had a mean rank of 2.614 ($p = .014$). Recall that for ranking, a lower number indicates greater preference and that students ranked the treatments in order of preference from 1 to 4. When pause conditions were analyzed in combination with the presence/absence of nth language, for conditions where nth language is not used, the treatment with uniform pauses was described as more helpful (1.5) than the treatment with graduated pauses (1.318, $p = .044$). Comparing rankings between these two treatments did not produce significant results.

Qualitative Findings

These were difficult expressions to process auditorily without navigation capability. The length and complexity of the expressions was deliberate, since we were exploring the usefulness of pauses and nth parenthesis language in assisting with understanding. We had originally considered using even longer, more complex expressions, but we decided that students would be likely to simply try to write the expressions down in order to work with them, and so we shortened them considerably in the hope that while they would still be sufficiently complex to show differential effects from different treatments, they would not be so complex as to overwhelm the students or to motivate them to write the expressions down in order to parse them. Nonetheless, many of the students simply found the expressions too much to handle without navigation or a visual or braille reference:

- It just seemed really long. It was too long for me to remember what was going on. If I was to solve it, I would have been lost.
- The way it was worded was fine, but if I had to solve it I would like to navigate it better.
- I don't think I could have done this type of math statement without an external refreshable braille display such as the 25 by 40 or the 40 cell line display. It is my preference that there be made an accommodation in the software settings for most refreshable braille displays.
- The wording could be helpful but I would need to see it to make sense. I am a visual learner.
- This was not helpful either because it would be easier to read it than to listen to it. It was a long expression.

Students who liked uniform pauses without nth language said about the example using that treatment:

- This one was more normal.
- I still like it better when they tell you when the parentheses open and close, and they are not numbered.

One specific objection to nth language said, "I mixed up the actual numbers with the parentheses numbers," and another student indicated that the numbering of parentheses interfered with their writing down the expression.

But more students expressed an appreciation for numbering the parentheses, especially when combined with uniform pauses:

- Not having pauses or numbers of parens doesn't help me keep track of things very easily.
- [2A (uniform pauses, no spoken nesting level) was] not helpful because it did not tell me if it was first, second or third paren.
- [2B (uniform pauses with spoken nesting level)] indicates the first second and third parenthesis but doesn't have any pauses or strange things to throw me off so I know exactly what everything is.
- The combination of the numbered sets of parenthesis and the faster pacing made this question [2B] the easiest to understand.
- Having first and second parens added is the main thing to understanding the math. The pacing and pauses aren't needed if you have that.
- The way it was read but I think having first and second parens is better than having the pauses [in 2C (graduated pauses, no spoken nesting level)].
- I found hearing the number for the parens helped me follow along [2D (graduated pauses, spoken nesting level)].

Some students liked the graduated pauses without the nth language:

- It was easier this time because not only did they have spaces, but the parentheses were less wordy. The spaces (pauses) helped decide which numbers were in which parentheses.

Others found the graduated pauses confusing or distracting:

- This time it ran together plus had random pauses [2C].
- When it uses first second and third parenthesis that is really helpful. But the pauses are a little confusing and I am not sure what the pauses mean. Then I get distracted trying to figure out what the pauses mean and forget the math [2D].
- The number of parens is good. Having both the number of parens and pauses too is a little too much [2D].
- The math statement hesitates quite a bit in the middle of the statement. If I were using the software now, and this were my math homework, halting speech with math phrases would not be my preference [2D].
- Not sure what the pauses do and what they mean … they were confusing [2D].
- I knew where the parentheses were starting, but not when they were stopping. Having the parentheses numbered wasn't as confusing this time [2D].
- Had random pauses and I didn't know when it was going say something [2D].
- I didn't like that it was slower. I didn't like the numbered parentheses. By the time the expression was over, I had forgotten what the beginning was. The numbered parentheses mixed with the numbers and the expression made it confusing to understand the problem [2D].

Discussion

Based on the quantitative analysis and on students' comments, some support was found for retaining uniform pauses and adding nth language to multiply-nested parentheses. Support was also found for interactive navigation as an aid to parsing long expressions.

Research Question 3 (Implied Times and Parentheses)

Whether to speak all parentheses or just those needed for clarity, and whether to speak "times" when parentheses are used to indicate multiplication of two quantities (e.g., $(x + 1)(x - 1)$), had been a subject of discussion among the project staff and consultants who were grappling with the larger question of the extent to which spoken math should replicate all symbols used in the print (or Nemeth Code) version as opposed to emphasizing the mathematical structure, regardless of how it is printed or brailled. This question underlies differences between the two competing approaches to math speech that we discuss in more detail elsewhere (Frankel et al., 2016): the classroom-like approach taken by ClearSpeak (derived from Chang's [1983] approach), or the Nemeth Code Braille-emulating approach taken by MathSpeak (Nemeth, 2013; gh, 2004–2015). If the expectation is that the spoken math would be used only or primarily to allow the student to transcribe it into print or braille (and then work primarily from the transcription), it is reasonable to speak the symbols. On the other hand, if the student is expected to work directly from the spoken math, speaking the symbols in addition to or

Table 4 Question 3 Treatments

Expression	Treatment Features		
	Pauses: M = Medium-length (400 msec) pauses before and after "times" was spoken; L = Long (800 msec) pauses before and after "times" was spoken	Parentheses Spoken	Speedup (125% of original speech rate) of parts of expression enclosed by parentheses
3A: $\frac{(3x-2)(x^2+4)(2x+3)(3x-2)}{(x^2+4)(2x+3)(3x+2)(2x-3)}$	M	Yes	No
3B: $\frac{(3x+5)(x^2+6)(5x-3)(x^2+6)}{(3x+5)(5x-3)(x^2-6)(5x+3)}$	M	Yes	Yes
3C: $\frac{(x+1)(x^2+4)(3x+7)(2x+5)}{(x^2-4)(3x+7)(x-1)(2x+5)}$	L	No	No
3D: $\frac{(x+1)(x^2-4)(5x+2)(2x+3)}{(2x+3)(x^2-4)(5x+2)(x-1)}$	L	No	Yes

Note. In all treatments, the implied "times" indicated by the parentheses was spoken as "times." If spoken, parentheses were spoken as "open paren . . . close paren."

instead of their mathematical meaning can create unnecessary memory load. Speaking certain symbols can also impede understanding for students who are familiar with the mathematical meanings but not with the symbols. To gain insight into this issue (and thus to guide our further development of ClearSpeak), we devised the questions in Section 3 to focus on parentheses and implied times, including the extent to which (in expressions of the sort tested) parentheses can remain unspoken when the implied times is spoken. Section 3 also looks at how pauses and/or speedups may impact intelligibility of the spoken expression, and so we designed an item intended to give insight into which of these prosodic and lexical cues are most useful to students. We created four treatments, each paired with an algebraic fraction. The different algebraic fractions were clones of one another in that they were similarly structured but not identical, so that knowledge of one could not be used to decode another one. All examples spoke the implied times. Table 4 shows the expressions and treatments.

For each of the four treatments (presented in varied orders), students were asked to identify, from a list of possibilities, which expression(s) were in both the numerator and the denominator. They were allowed to listen to each expression as many times as they wished. The question for Expression 3A was as follows (the questions for the other expressions were similarly structured, but the answer choices and correct answer were specific to each expression):

> Which expressions are in both the numerator and the denominator)? There could be more than one correct answer. If so, type all of them.
>
> a. $3x$ plus 2
> b. $3x$ minus 2
> c. $2x$ plus 3
> d. $2x$ minus 3
> e. $2x$ plus 2
> f. $2x$ minus 2
> g. x squared plus 4
> h. I can't tell

Table 5 Question 3 Results

Treatment	3A: Parens spoken; 400 msec pauses, no speedup	3B: Parens spoken; 800 msec pauses and speedup (to 125% of original speech rate)	3C: Parens not spoken; 800 msec pauses, no speedup	3D: Parens not spoken, 800 msec pauses and speedup (to 125% of original speech rate)
Math correctness	Ave: 0.81	Ave: 0.81	Ave: 0.88	Ave: 0.82
	SD: 0.168	*SD*: .247)	*SD*: 0.172	*SD*: 0.149
How sure (0–3)	Ave: 2.27	Ave: 2.24	Ave: 2.32	Ave: 2.23
	SD: 0.767	*SD*: 0.994	*SD*: 0.716	*SD*: 0.869
Helpfulness (0–3)	Ave: 1.82	Ave: 1.86	Ave: 1.73	Ave: 1.77
	SD: 0.958	*SD*: 0.99	*SD*: 0.883	*SD*: 1.066

This question (and the corresponding questions for the other expressions in this section) was scored based on how many of the seven possibilities a student correctly identified as appearing (or not appearing) in both the numerator and the denominator. In this instance, the correct response was choices *c* and *g*, so a student who selected *c*, *g*, and nothing else correctly identified seven of the seven possibilities and was assigned a score of 1 for that item. A student who selected *d* and *g* correctly identified five of the seven (incorrectly identified *d* and incorrectly failed to identify *c*), and so received a score of 5/7 or 0.714 for that item.

Following the math question, students were asked the same types of feedback questions as for the other sections (how sure they were of their answer, how helpful the speech was, and the open-ended question about what aspects of the speech helped or made it difficult to answer the math question).

Question 3 Results

Quantitative Analysis

As can be seen from Table 5, the results were very similar for all conditions. Statistical analysis found no statistical significance in any of these comparisons.

Qualitative Findings

Consistent with the lack of significance found by the quantitative analysis for Question 3, analysis of students' comments showed that students were similarly divided in their expressed preferences for each of the four treatments. For example, regarding Statement 3C, which omitted speaking open and close parentheses, replacing them with a longer pause around the "times," one student said, "Better than the last time. Read it slower paced, which was better," while another said, "The slower pace was harder to follow." One missed hearing the parentheses: "It made it extremely difficult that it did not specify where the parentheses were, and based on the questions, I knew there should be parentheses," while another thought they were extraneous: "Not having open and close was better because that isn't really needed and just adds extra stuff." On statements using "paren," one student said, "All the open parens and closed parens made it harder to remember what it said." Similarly, some students found it helpful to hear "times" and others found it got in the way.

Many students commented on how easy or difficult the expression was to write down, based on the speed.

- "The time to say everything took too long for me to remember everything in my head."

Some requested navigation:

- "It made it difficult by how quickly it said it. I couldn't catch everything it was saying unless I was writing it down. I needed to be able to hear it piece by piece. I think you should add that feature."
- "There should be a way to navigate through the problem instead of having to listen to the whole thing over and over again pushing the Ctrl key to stop after each chunk."

- "Too difficult to copy. It talked too fast for me to copy, and the numerator and denominator are not separate chunks or pieces so when copying it makes it difficult. I would like for you to be able to arrow up and have it read one section and arrow down to read another section."

Discussion

With the types of lengthy expressions used in Section 3, differences in wording and prosody seem to have little effect on students' success at decoding the expression. Without the ability to navigate the expressions, students just wanted to write them down and work with them that way. So although this section of the study did not do very much to advance our thinking on prosody, it did support the need for navigation, which was implemented and tested in the third feedback study and pilot.

Conclusions

As noted in our discussion of Research Question 1, this study provided support for end language to delimit the scope of radicals and, perhaps by extension, similar language for other structures where the end of the structure is clarified by its insertion (exponents, absolute values, fractions), and less support for uses of extended pauses or speeded-up speech to signal scope. However, it left open the possibility that with increased familiarization or instruction, the enhanced prosodic elements tested, alone or combined with end language, might become more helpful for some types of expressions or circumstances. Further research might yield useful information in that regard.

We noted in our discussion of the Research Question 2 that the study provided some support for using language ("nth parenthesis") to help track the levels of nested parentheses. There did not appear to be support (at least without additional familiarization) for using longer or graduated-length pauses or for using speeded-up speech for signaling what was within parentheses or how deeply parentheses were nested.

The results of testing Research Question 3 did not provide any clear direction for prosody. It suggested that when listening to a lengthy expression, such as those used for the third question, students rely heavily on writing the expression down and feel that navigation capability would be useful. It is reasonable to expect that the more complicated the expression is, the more difficult it is to communicate its contents with a single audio read-through (or even multiple audio read-throughs). Although prosody may help in such cases, navigation is necessary to understanding it purely through audio, and a full read-through of an expression in context may serve primarily to provide a rough idea of where the expression fits in with its context.

Note that although we tested specific prosodic enhancements in this study, some prosody had already been built in to ClearSpeak (as mentioned previously) and remains a component. The baseline ClearSpeak prosody includes small pauses that make the computer speech sound more natural and less rushed. That is, we did not test purely "flat" speech; the degree of help provided by ClearSpeak's baseline prosody could potentially be measured by a study comparing it with a version of ClearSpeak from which all prosodic elements had been removed (but retaining its lexical rules and preferences). However, as suggested by Gellenbeck and Stefik (2009), purely flat speech is likely to be found less useful than speech with natural pauses.

Overall, then, the study suggests a case for some combination of increased familiarization with some of the enhanced prosodic elements and for interactive navigation. The latter, as noted in the "Background" section, was recommended by some other authors and allows listeners to control the pace and granularity of the math they are listening to. Interactive navigation was, in fact, developed in the next phase of this project and the results are reported in Frankel, Brownstein, and Soiffer (in press).

Another avenue for further research is to investigate prosody in more detail. For example, although our prosodic cues were based loosely on those observed in human speech, more detailed investigations could be conducted to examine variations and commonalities among human-spoken versions of sample expressions. Another potential area of study would be a direct comparison of the usefulness of prosody (especially where explanations and practice are provided prior to testing) with that of interactive navigation.

Acknowledgments

We are grateful to Susan Osterhaus and Maylene Bird of the Texas School for the Blind and Visually Impaired and to Stephen L. Noble, Research Consultant, University of Louisville and Content Specialist: Mathematics and Science, Bridge MultiMedia for their assistance with the development of the research instruments and with administering the instruments to participating students. Staff at the Washington State School for the Blind also recruited students and administered the instruments to them. Neil Soiffer, of Design Science, Inc., led the development of the software that enabled ClearSpeak to exist and enabled us to experiment with the speech elements described in this report. G. W. Micro enhanced its Window-Eyes screen reader to support MathPlayer, including the prosodic elements available in ClearSpeak inside Microsoft Word as well as in web browsers. Nan Kong, with consultation from Eric Hansen, performed the statistical data analysis.

The research reported here was supported by the Institute of Education Sciences, US Department of Education, through Grant R324A110355 to the Educational Testing Service. The opinions expressed are those of the authors and do not represent views of the Institute of Education Sciences or the U.S. Department of Education.

Note

1 While our project was in progress, some screen reader support for reading and navigating math expressions presented in MathML was developed for ChromeVox, VoiceOver, and JAWS. See Soiffer, Frankel, and Brownstein (2015) for a video demonstration of the differences in approach. Note that unlike our project, the supports in the environments just noted work only in certain browsers and not at all in Microsoft Word.

References

Bates, E., & Fitzpatrick, D. (2010). Spoken mathematics using prosody, earcons and spearcons. In K. Miesenberger, J. Klaus, W. Zagler, & A. Karshmer (Eds.), *Proceedings of the 12th International Conference on Computers Helping People With Special Needs* (pp. 407–414). Berlin, Germany: Springer-Verlag.

Blackorby, J., Chorost, M., Garza, N., & Guzman, A. (2003). The academic performance of secondary school students with disabilities. In M. Wagner, C. Marder, J. Blackorby, R. Cameto, L. Newman, P. Levine, & E. Davies-Mercier (Eds.), *The achievement of youth with disabilities during secondary school*. Retrieved from http://www.nlts2.org/reports/2003_11/nlts2_report_2003_11_ch4.pdf

Ferreira, H., & Freitas, D. (2005, February). *AudioMath—Using MathML for speaking mathematics.* Paper presented at the meeting of XML: Aplicacoes e Tecnologias Associadas, Braga, Portugal.

Frankel, L., Brownstein, B., & Soiffer, N. (in press). *Expanding audio access to mathematics expressions by students with visual impairments via MathML.* Princeton, NJ: Educational Testing Service.

Frankel, L., Brownstein, B., Soiffer, N., & Hansen, E. (2016). *Development and initial evaluation of the ClearSpeak style for automated speaking of algebra* (Research Report No. RR-16-23). Princeton, NJ: Educational Testing Service. http://dx.doi.org/10.1002/ets2.12103

Gellenbeck, E., & Stefik, A. (2009). Evaluating prosodic cues as a means to disambiguate algebraic expressions: An empirical study. In S. Trewin & K. R. McCoy (Eds.), *Proceedings of the 11th International ACM SIGACCESS Conference on Computers and Accessibility* (pp. 139–146). New York, NY: ACM.

gh. (2004–2015). *MathSpeak grammar rules.* Retrieved from http://www.gh-mathspeak.com/examples/grammar-rules/

Karshmer, A., Gupta, G., & Pontelli, E. (2007). *Mathematics and accessibility: A survey* (University of Texas at Dallas Technical Report). https://www.utdallas.edu/˜gupta/mathaccsurvey.pdf

Nemeth, A. (1972). *Nemeth braille code for mathematics and science notation.* Frankfort, KY: American Printing House for the Blind.

Nemeth, A. (2013). *MathSpeak.* Retrieved from http://accessinghigherground.org/handouts2013/HTCTU%20Alt%20Format%20Manuals/Math%20Accommodations/07%20MATHSPEAK.pdf

Raman, T. V. (1994). *Audio system for technical readings.* Ithaca, NY: Cornell University.

Soiffer, N., Frankel, L., & Brownstein, B. (2015, March). *Survey of math accessibility.* Power point presented at the annual meeting of the International Technology and Persons with Disabilities Conference, California State University, Northridge. Retrieved from http://www.dessci.com/en/reference/ies-ets/conferences/csun_2015/Survey%20of%20Math%20Accessibility.pptx

Stevens, R., Edwards, A., & Harling, P. (1997). Access to mathematics for visually disabled students through multi-modal interaction. In *Human–computer interaction* (Vol. 12, pp. 47–92). Mahwah, NJ: Lawrence Erlbaum Associates.

Appendix A

For a version of this report that is fully accessible using the tools described, download the Microsoft Word document located at https://www.ets.org/Media/Research/RR-16-33.docx. Accessible reading also requires:

MathPlayer: http://www.dessci.com/en/products/mathplayer/download.htm

MathType: http://www.dessci.com/en/products/mathtype/default.htm

MathPlayer is free; MathType is a paid product, but is available for free trial.

If screen reader integration is desired, download the free NVDA screen reader: http://www.nvaccess.org/download/.

Additional tools, tutorials, and related information can be found at http://www.clearspeak.org.

Suggested citation:

Frankel, L., & Brownstein, B. (2016). *An evaluation of the usefulness of prosodic and lexical cues for understanding synthesized speech of mathematics* (Research Report No. RR-16-33). Princeton, NJ: Educational Testing Service. http://dx.doi.org/10.1002/ets2.12119

Action Editor: Keelan Evanini

Reviewers: James Fife and Ruth Loew